IBM

Nestlé achieves Global Excellence with SAP NetWeaver BI and IBM

"The proof of concept performed with IBM and SAP has been of real value to Nestlé. It has demonstrated that it is possible to scale to the very large volumes that we expect to reach. The fact that the test combined the skills of IBM, SAP and Nestlé and included hardware, database and applications gives us a lot of confidence in the results. Secondly, it has enabled us to establish the target architecture we believe is required and to have confidence in that architecture. Thirdly, because of the testing carried out at various stages, it has also helped us in determining our road map and steps for moving to our target architecture."

Chris Wright
Nestlé GLOBE Business Warehouse Manager
Nestlé

Nestlé achieves Global Excellence with SAP NetWeaver BI and IBM

About this paper

This technical brief describes a proof of concept for the implementation of a 60TB SAP® NetWeaver Business Intelligence (SAP NetWeaver BI) system on behalf of the Nestlé worldwide Global Excellence project. In the Nestlé GLOBE project, the SAP NetWeaver BI system plays a strategic role as it is the focal point for business process data over multiple markets and geographies, and provides the consistent view necessary for global decision making.

This proof of concept was completed jointly by SAP and IBM to prove to Nestlé that a system of the magnitude proposed is feasible, can be administered with current tools, and can provide the performance Nestlé requires for the rollout of the GLOBE project.

Customer Objectives

● Implement a consistent end-to-end business view over the business systems and processes in multiple geographies, by consolidating within SAP BI multiple sources of business data such as manufacturing, sales, logistics, finance

● Provide an effective, high-quality tool with near real-time information, able to support executive business decision making, through comparisons of actual and historical data

● Improve decision quality by enabling current positioning with historical data and tracking trends in a global context.

Resulting Customer Requirements

● Extremely large database requirement expected to exceed 60TB

● Ability to maintain and protect 60TB of data with limited disruption to business

● Quick reporting response times for "ad hoc" as well as standard queries

● Quick deployment of new data into the system: "near real time"

● Support a combination of the previous workload types.

IBM Solution

● Use the benefits of DB2 Database Partitioning Function (DPF) to manage high-end scalability and provide very large data access bandwidth

● Implement Tivoli Storage Manager and Tivoli Storage Manager for Advanced Copy Services (ACS) for parallel backup/restore and disaster recovery to compress maintenance windows and recover time

● Use IBM POWER5 virtualization functionality to support volatile load profiles and to support both production load and concurrent administration requirements

● Provide high capacity I/O performance using IBM DS8300 with IBM FlashCopy functionality to support both performance requirements and concurrent administration needs.

"For industry leaders like Nestlé, the rapid delivery of business intelligence is becoming a critical part of operations. This outstanding Proof-of-Concept has helped Nestlé gain a more strategic view of its business data – and demonstrates that IBM solutions can deliver the levels of performance and scalability required for even the largest and fastest-growing SAP NetWeaver BI environments."

Volker Loehr
General Manager, Global IBM SAP Alliance
IBM

Background, starting point and objectives

Nestlé strategy

Nestlé is the largest worldwide company in the food and beverages sector, with sales for 2006 up to CHF 98.5 billion and a net profit of CHF 9 billion. Headquartered in Vevey, Switzerland, the group employs more than 260,000 employees and has factories or logistics operations in almost every country. In recent years, Nestlé launched the world's largest SAP implementation project called Nestlé GLOBE, designed to unlock Nestlé's business potential by creating and adopting common global business processes while allowing each market to make their specific analysis and take local decisions.

Implementing and running a business intelligence system supports the Nestlé business strategy. This repository of consolidated data combining production data extracted from the day-to-day operations as well as sales and financial data representing the market trends is considered a strategic tool for informed business decisions and further business development.

For international companies such as Nestlé that run businesses all over the world, these consolidated data warehouses can quickly become extremely large. Positioned at the heart of their strategy, the reliability, scalability and manageability of this system's underlying solution is vital in every sense. Data Management is a critical aspect of the whole system as Nestlé is forecasting that strategic data consolidation will increase significantly in the next few years.

Starting Point at Nestlé

Nestlé had already decided to implement IBM DB2 Universal Database, using DPF (Database Partitioning Function). The Nestlé team was convinced that this partitioned database was the most likely technology to be capable of scaling to 60TB and beyond.

Nestlé had already begun to implement a limited DB2 DPF in its SAP NetWeaver BI production environment of 7TB (status December 2005). The production hardware configuration, an IBM System p model p5 595 and a single IBM System Storage model DS8300 storage server were considered capable of scaling to 20TB, the expected data volume by December 2006. At this point new hardware would be required and Nestlé needed a proven design for both hardware and database, and the certainty that the proposed infrastructure and the SAP applications were capable of meeting the business requirements at 60TB.

IBM & SAP working together

Beginning in December 2005, IBM and SAP engaged in cooperation to perform a proof of concept for the next generation of very high-end business intelligence requirements.

Using the infrastructure building blocks already successfully implemented by Nestlé – IBM p5 p595, IBM DS8300 Storage, DB2 DPF, and Tivoli – IBM and SAP proposed to demonstrate the following:

- Optimal DB2 DPF design for high scalability
- Optimal storage server design for performance and administration of extremely large databases
- Optimal Tivoli design for highly parallel administration of extremely large databases
- Proof of high-end scalability for the SAP NetWeaver BI application
- Proof of infrastructure and application design for "high-end" performance scalability
- Proof of infrastructure flexibility supporting a broad mix of workload types.

The objective was to prove that by using a solution based on massive parallelism, that extremely large databases can be managed using tools available today, and can be managed in the maintenance windows typical for business-critical production systems. This was a proof of concept for the largest SAP database build to date, and it was to prove that the infrastructure and application could be designed to achieve extreme workload scalability and variation. This proof of concept was a "first of a kind:" it had no precedence at any component level.

The proof of concept would chart the future direction that Nestlé's infrastructure should take in order to support the predicted growth of the BI system in terms of size and load. The key performance indicators (KPIs) of the proof of concept were modeled to fulfill the expectations of the Nestlé workload as it is expected to evolve over the next two to three years, and based on actual business processes and design criteria. The challenge was to demonstrate generally the scalability, manageability, and reliability of the BI application and the infrastructure design, as BI requirements begin to necessitate multi-terabyte databases.

"The major challenges we had to overcome in terms of architecture were the selection of the right component characteristics to both deliver the end-to-end bandwidth to access the 60TB of active data, and to support the broad mix of workload types. Thanks to the high-end technology, combined into the largest DB2 DPF on System p and DS8300 that has ever been designed and tested, and thanks to our unmatched team expertise, we successfully delivered a 'near real-time' SAP Business Intelligence system that should help to meet Nestlé's predicted requirements for two to three years from the start of the project."

Francois Briant
Executive IT Architect and SAP BI PoC Lead Architect
IBM

The Proof of Concept

Key Performance Indicators – the challenge

The proof of concept was based on KPIs identified by Nestlé as critical to support the business. These KPIs are divided into two categories: administration and performance.

Administration KPIs

The administration KPIs are those which the infrastructure must achieve to prove the maintainability of the database and disaster recovery approach. The major KPIs are documented in the table.

This proof of concept reproduced the BI system growth from 7TB to 20TB and finally to 60TB. At each of these strategic phases of data growth, a series of tests were performed for verification of the KPIs for administration and performance. Infrastructure KPIs had to meet "strict targets set by the business around the maintenance time window", even as the database increased to nearly nine times the baseline size.

Combined performance KPIs

The performance KPIs focused on the time-critical application processes, including the high-volume integration of new source data as well as online reporting performance. The tests simulated a 24x7 Service Level Agreement over multiple

Ref.	Description	Requirement
KPI-1	Recovery of an entire days work Restore Database and roll forward 500GB of logs	<8hrs
KPI-2	Disaster Recovery Complete Database restore from tape + roll forward of 2TB of logs	<24hr
KPI-3.a	Daily Backup to Tape Full Database backup to Tape	<12hr
KPI-3.b	Daily Backup to Disk Completed processing of incremental FlashCopy of production day	<8hr

Figure 2 - 1: Administration KPI

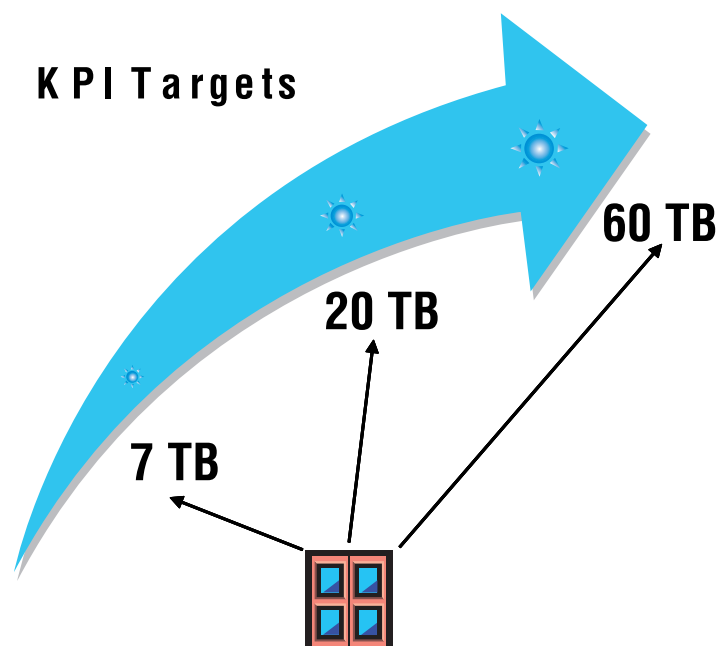**K P I T a r g e t s**

60 TB

20 TB

7 TB

Figure 2 - 2: Database sizes

geographies, a business environment common to many large enterprises.

These tests represented a combined load scenario, Figure 2 - 3, in which data translation or cube loading (2), aggregate building (3), and online reporting (4) with fixed response time requirements, all ran simultaneously. This combined load increased over the lifetime of the project to five times the peak volume observed on the Nestle Production System at the start of the project.

The query design for the proof of concept was implemented to reproduce behavior observed in the real production system. The reporting queries ranged over 50 different reporting cubes, and 50% of the queries could take advantage of the application server OLAP cache, while 50% could not. Of the queries, 80% used the BI aggregates and 20% went directly to the fact tables, and these latter 20% were the major challenge. The database load, from these reports using the fact tables, was also affected as the database grew and the number of rows in the fact tables increased from 20 to 200 million rows.
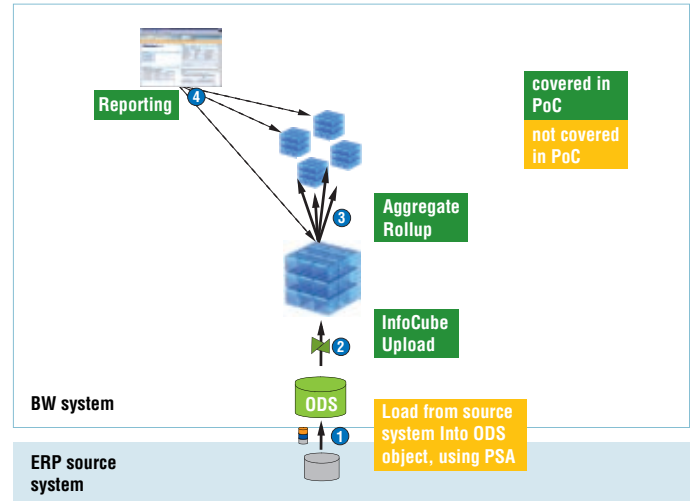
Figure 2 - 3: the 3 critical online load requirements
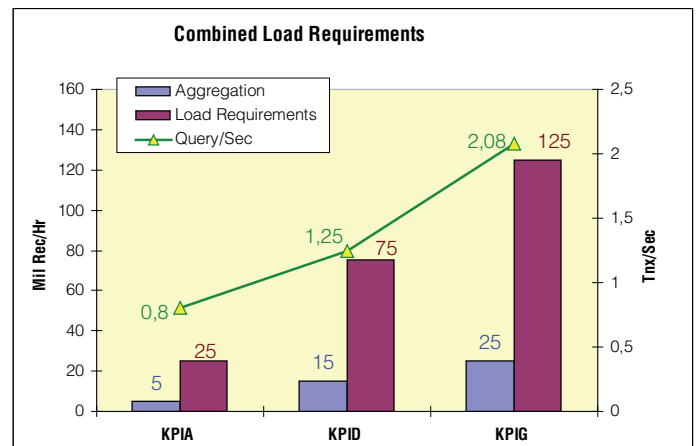
Figure 2 - 4: Graph online KPI scale up to simulate expected rollout – Factor 5
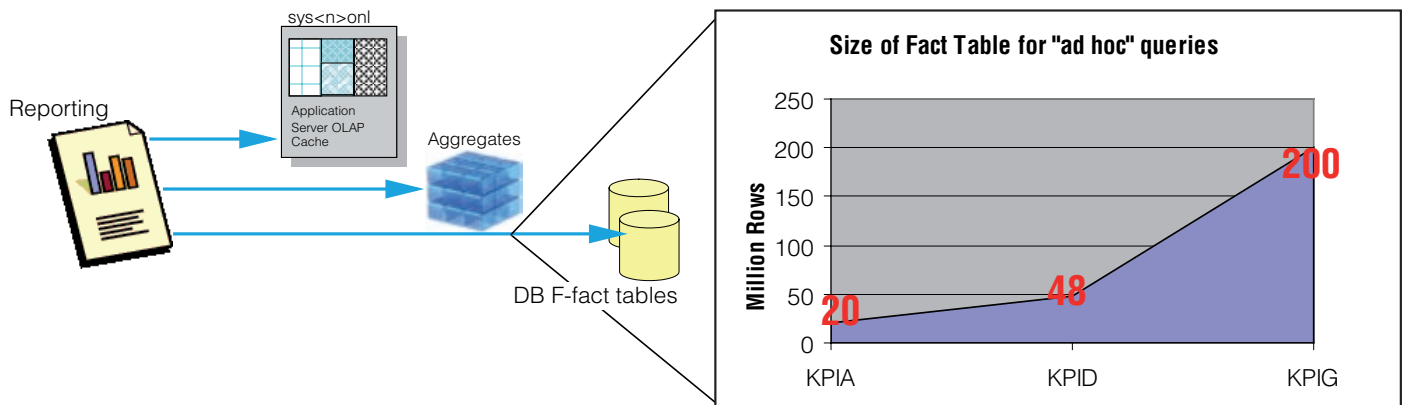

Figure 2- 5: Report simulation and data growth of fact-table target

Solution architecture

DB2 database design decisions

Working together with the DB2 development lab in Toronto, Canada, a DB2 design for 33 database partitions was selected. The database would consist of one database server containing only partition 0, and 4 additional servers housing 32 additional database partitions (8 per server). This design was expected to provide the flexibility required to implement the 60TB database.

In the finalized design, each DB2 partition had dedicated file-systems and dedicated disks (LUNs). This design provides the most flexibility should it be necessary to redistribute the database partitions. The layout design of eight DB2 partitions per LPAR proved to be very effective, requiring no further redistribution in the proof of concept. The dedicated I/O design also has advantages for error and/or performance analysis in real life, as it is easier to trace the workload to LUN.

Partition 0 has a unique role in DB2 DPF; it maintains all the client communication connections, performs any necessary data merging of results data coming from the parallel partitions, and performs various administration tasks.

This partition was also used for less volatile data such as the dimension tables. The fact tables and aggregates were distributed across the parallel partitions.

According to size, the tablespaces containing active data were spread across eight to twelve DB2 partitions. Care was taken that these partitions were also equally spread across the physical server hardware. Using the data distribution design described above, a very even data distribution was achieved over the 32 DB2 partitions used for production data, delivering the extreme bandwidth required to access the very large data tables within the defined performance targets. Figure 2 - 8 depicts the final data growth stage at 65TB total data.
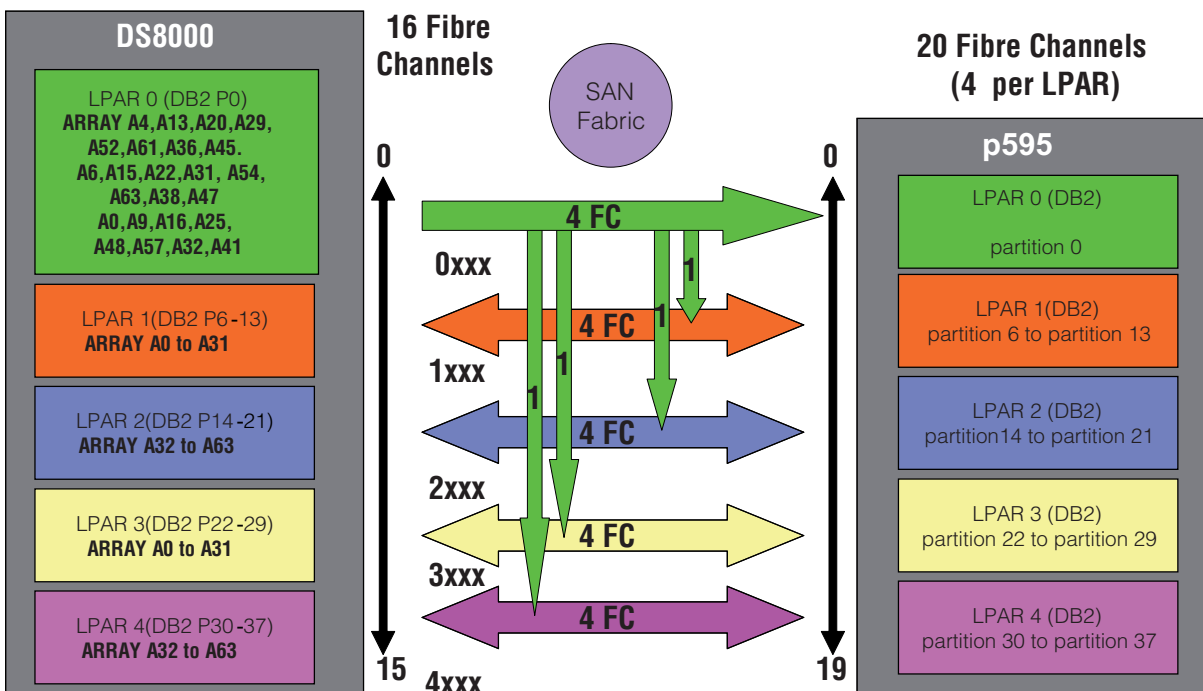

Figure 2 - 6: 33 DB2 DPF partitions with shared nothing infrastructure – Phase 1, 20TB
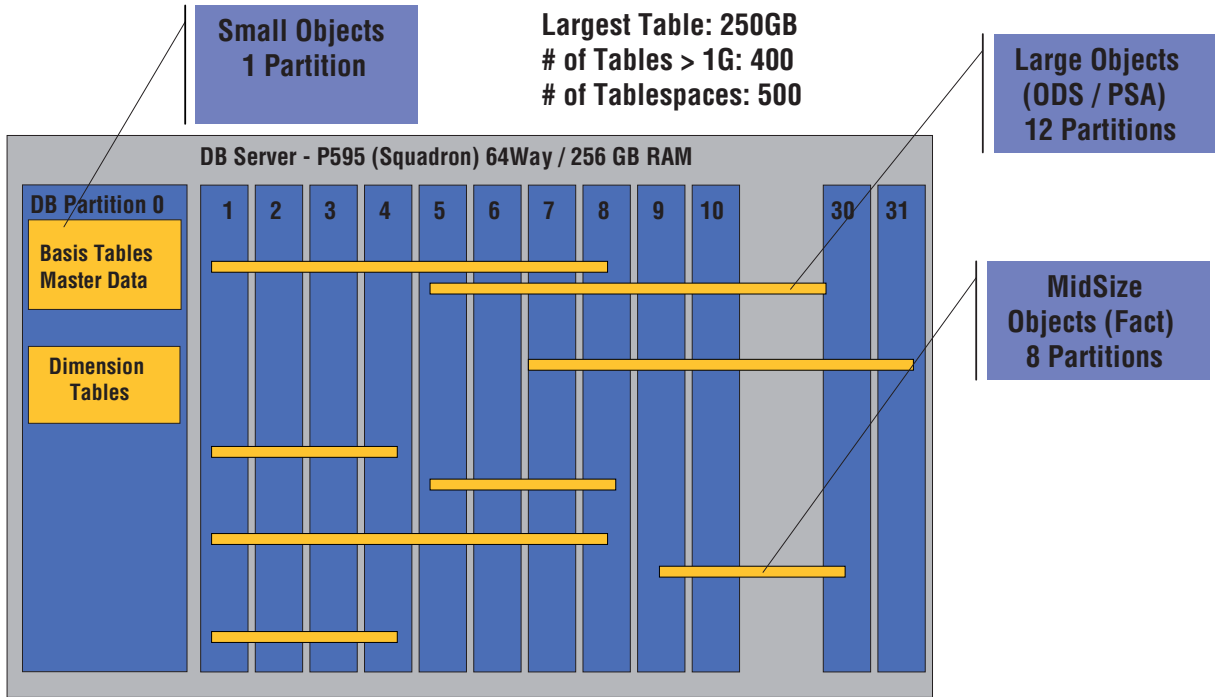
Small Objects
1 Partition

Largest Table: 250GB
# of Tables > 1G: 400
# of Tablespaces: 500

Large Objects
(ODS / PSA)
12 Partitions

DB Server - P595 (Squadron) 64Way / 256 GB RAM

DB Partition 0

Basis Tables
Master Data

Dimension
Tables

MidSize
Objects (Fact)
8 Partitions

1 2 3 4 5 6 7 8 9 10 30 31

Figure 2 - 7: DB2 DPF data distribution design

File Sytems Space allocated for the 60 TB database
Total = 65 TB
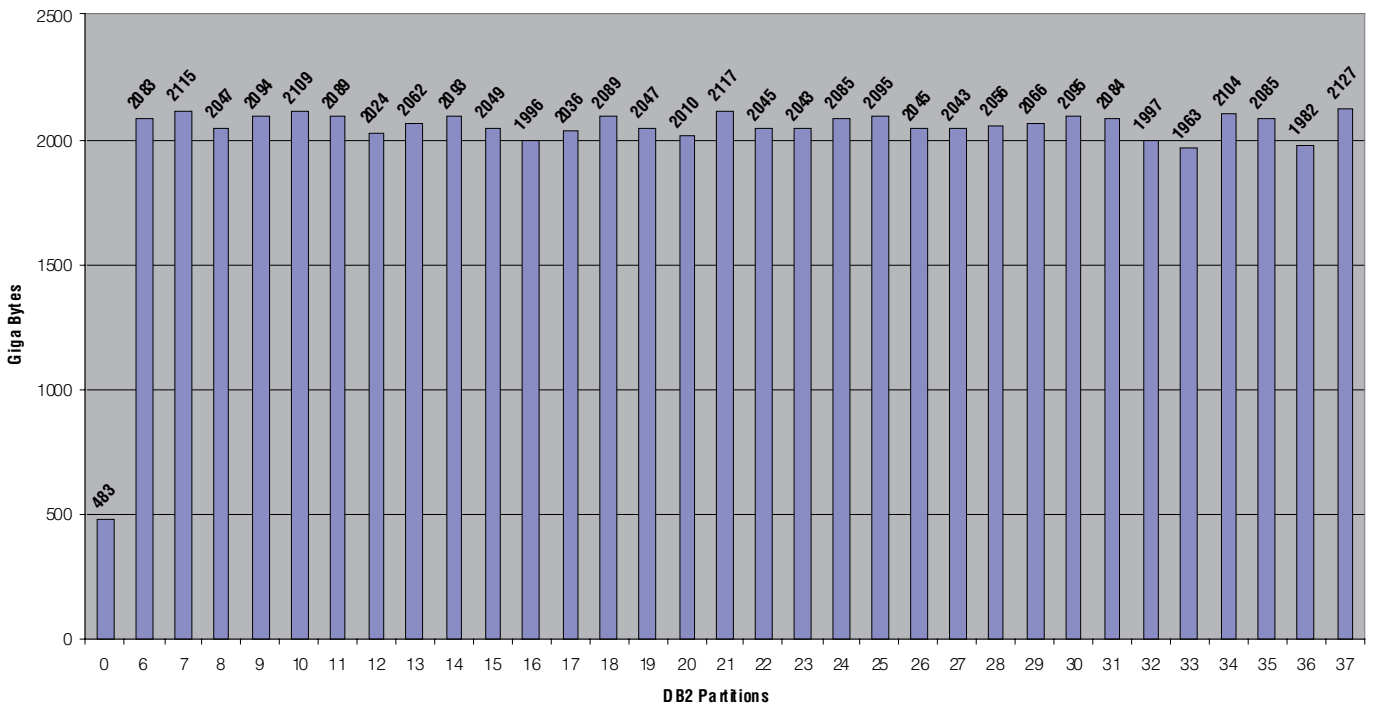
Figure 2 - 8: Data distribution over all 33 DPF partitions at 65TB

Evolution of DB2 in the proof of concept

The database used for the proof of concept was an actual clone of the Nestlé system. At the time it was cloned, DPF was already being used, but only in a small configuration.

There were six active DB2 partitions housed in a single database server with shared file-systems. The first steps in the proof of concept were to implement a new storage design with dedicated file system and storage per DB partition, and redistribute the database over five servers, which are logical partitions on a p595 server.

In Phase 1, the database was limited (by Nestlé specification) to a single p595 to reflect the current Nestlé hardware landscape. Phase 1 proved that the current landscape could scale to 20TB by using an optimal database and storage design.

The focus of Phase 2 was scalability of the design to 60TB. Phase 1 proved that the DB2 design, of eight partitions per LPAR, scaled well and this design was taken into phase2. The LPARs were distributed over five p595 servers to allow for the necessary memory increase required by the larger database, and to take advantage of the p595 resource virtualization. In the final layout, the database shares p595 CPU resources with the application servers, the Tivoli Storage Manager agents and the administration servers.

DS8300 layout design decisions

There were two major design decisions for the storage layout, and although there are many successful high-performance storage implementations in production, in this "first of this kind" new criteria had to be considered.
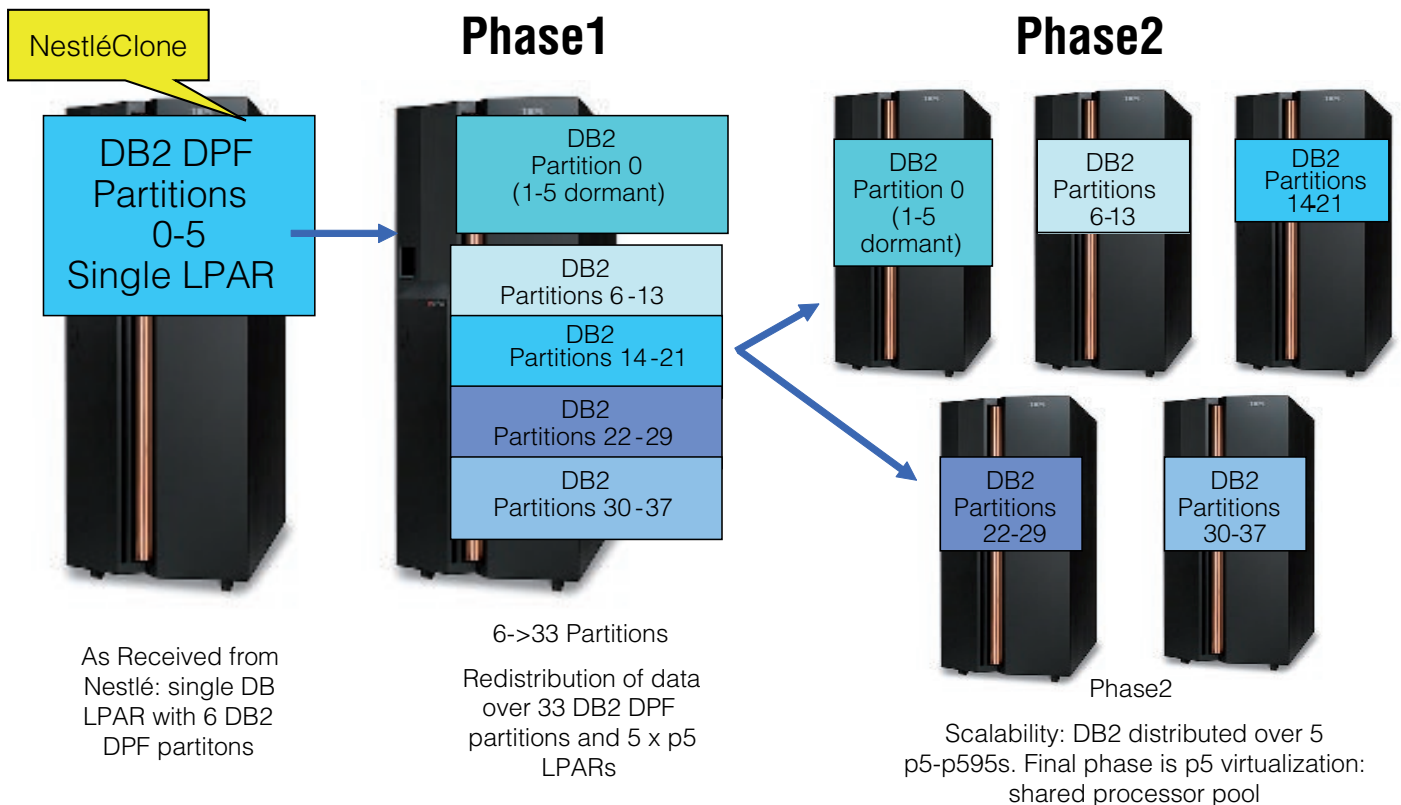

Figure 2 - 9: Evolution of the database server during the proof of concept

Data distribution options

a) Spread all the DB2 partitions on all the arrays by using small LUNs and having one LUN per array in each DB2 partition

b) Dedicate a group of arrays for each DB2 partition by using big LUNs in the group of dedicated arrays.

In a parallel database, where the data is spread very equally across all the parallel partitions, it is assumed that the I/O activity for each partition will be similar and that the load will be concurrent. By using dedicated storage arrays, the likelihood of I/O contention is reduced. Dedicated arrays allow the use of larger LUNs as the data does not need to be distributed over as many arrays, and fewer LUNs are needed. Using fewer LUNs is advantageous to the backup/restore and failover scenarios as there is some administration overhead per LUN.

Positioning of source and target LUNs for FlashCopy options

c) Dedicated arrays for FlashCopy LUNs: half the arrays for source, half for target

d) Spread source and target LUNs over all arrays

By separating the target and source LUNs on dedicated arrays, we can ensure that there will be no influence on the production disks resulting from activity on the FlashCopy targets used for backup. At the same time however, the number of spindles available for production I/O load is reduced. As the FlashCopy activity was not expected to have a major influence on performance, the decision was to place both target and source on all ranks, but ensuring that source and target pairs were placed on different ranks. This design allowed all spindles to be available for production requirements.
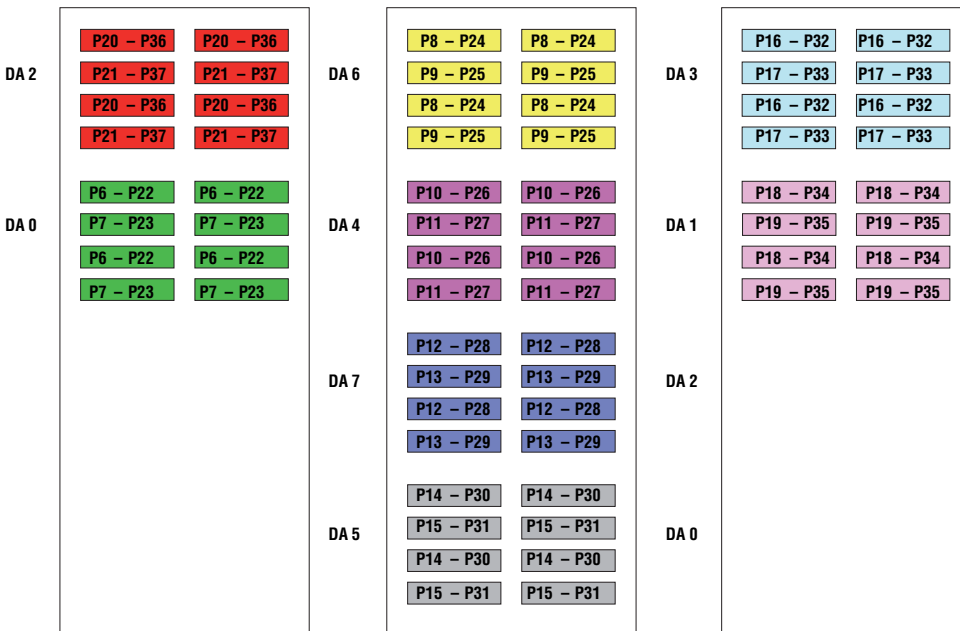

Figure 2 - 10: Dedicated storage ranks in Phase 1 – 20TB

Figure 2 - 11: Source and target pairs on alternate ranks

Evolution of the storage server during the proof of concept

Phase 1 was limited to a single p595 database server and a single DS8300 storage server to achieve scalability to 20TB. The first step in the proof of concept was to implement the storage design discussed above: dedicated storage arrays, dedicated LUNs, dedicated file-systems. The storage server was upgraded to the newest technology, DS8300 Turbo (2.3GHz) for the KPI-D load scaling stage. Phase 1 proved the performance of the single storage server to 20TB and three times the current Nestlé high-load requirements.

The Phase 1 design was carried over to Phase 2, distributed over four DS8300 Turbo storage servers, for the growth to 6TB.

"This Proof of Concept demonstrates the strength of the two companies combined technologies and skills, and the mutual commitment to excellence. We brought together a team of 25 experts, spanning 5 countries over two continents, for more than 14 months."

Herve Sabrie
Server Specialist and Nestlé SAP BI PoC Project Manager
IBM

Phase1

Phase2

Redistribution of DB2 data from shared to non-shared filesystems

Grow to 20TB

DS8300
7-20TB

DS8300Turbo
20TB

Redistibute the DB2 nodes over 5 LPARs and 4 storage servers

Grow to 60 Terabyte

Ea. 512 Disks (146GB 15K rpm) 54 TBs Capacity (production + FC)

4 x DS8300Turbo
60TB
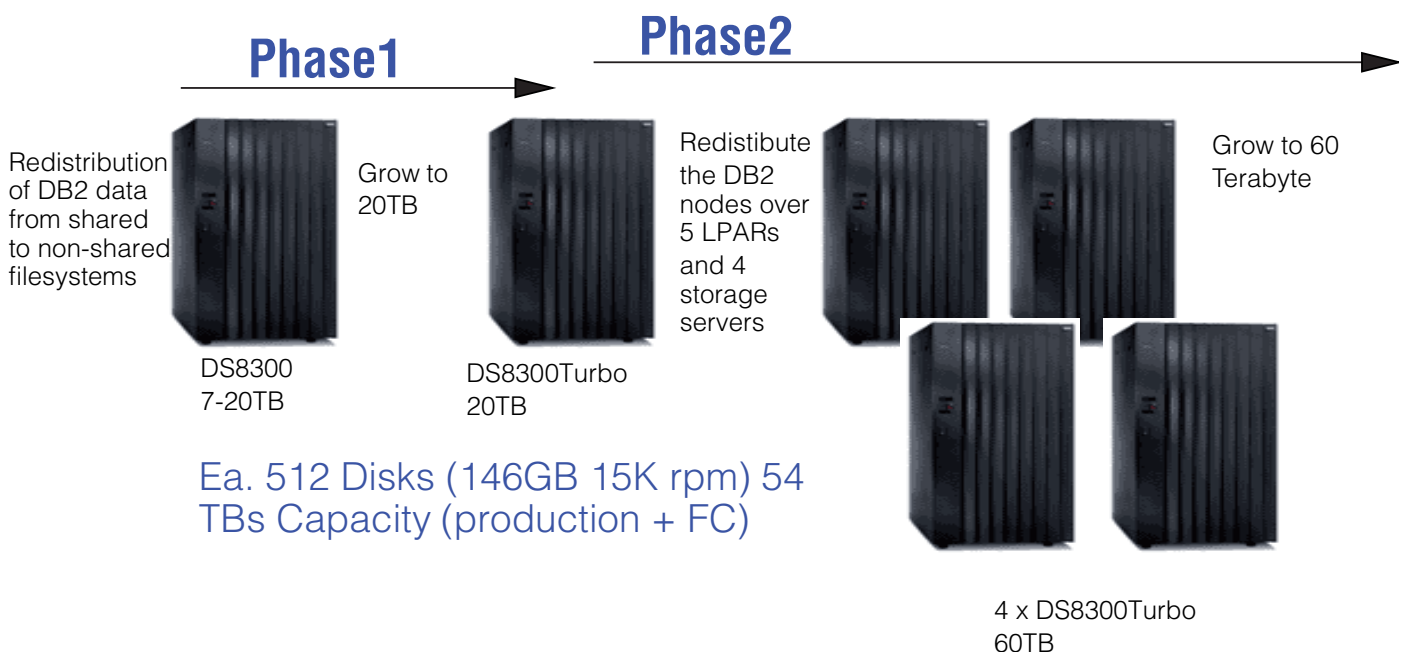
Figure 2 - 12: Evolution path DS8300

Proof of design

Figure 2 - 13 shows the CPU distribution within one database server with eight DB2 partitions. Each partition is represented here by a Workload Manager (WLM) class which monitors its CPU utilization. All partitions are equally active and the load is well distributed.

These two graphs in Figure 2 - 14 show the I/O behavior of two DB2 partitions over the same time frame. The two individual DB2 partitions show similar I/O behavior and peak load requirements. This demonstrates a good database load distribution and also shows that the disk layout decision based on dedicated LUNs was correct. I/O contention, caused by the simultaneous I/O activity is avoided, and the full benefit of the parallel database can be realized.
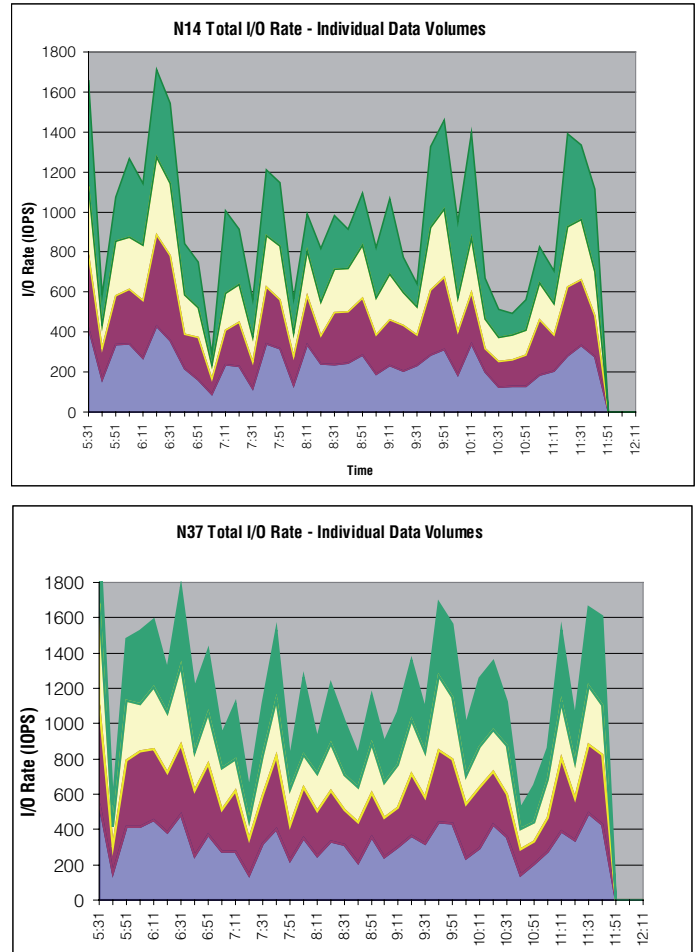
Figure 2 - 14 : I/O performance on two separate DB2 partitions
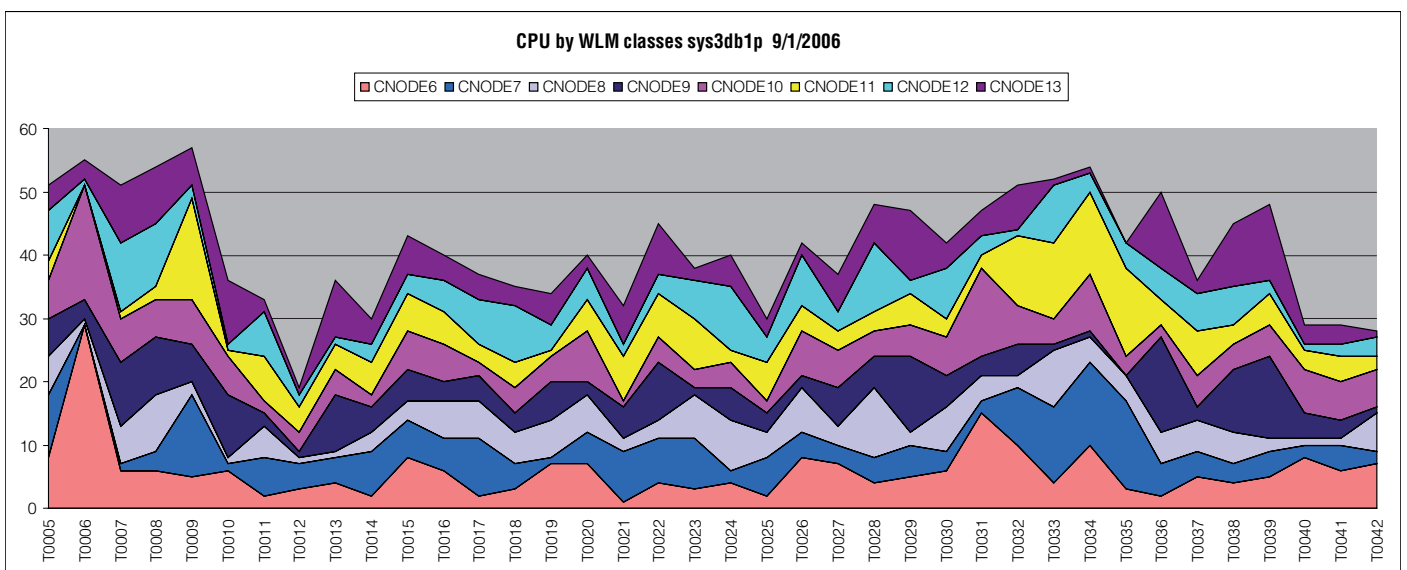
Figure 2 - 13: Load distribution over eight DB2 partitions in a p595 LPAR

Backup/Restore architecture

The Tivoli design implements five storage agents for better CPU utilization via cross-LPAR load balancing. As partition 0 must be backed up as the first step, four storage agents would have been sufficient. Four CIM agents were also implemented for faster and more reliable execution of FlashCopy commands. The storage agents and CIM agents were implemented on AIX 5.3 in shared processor LPARs.
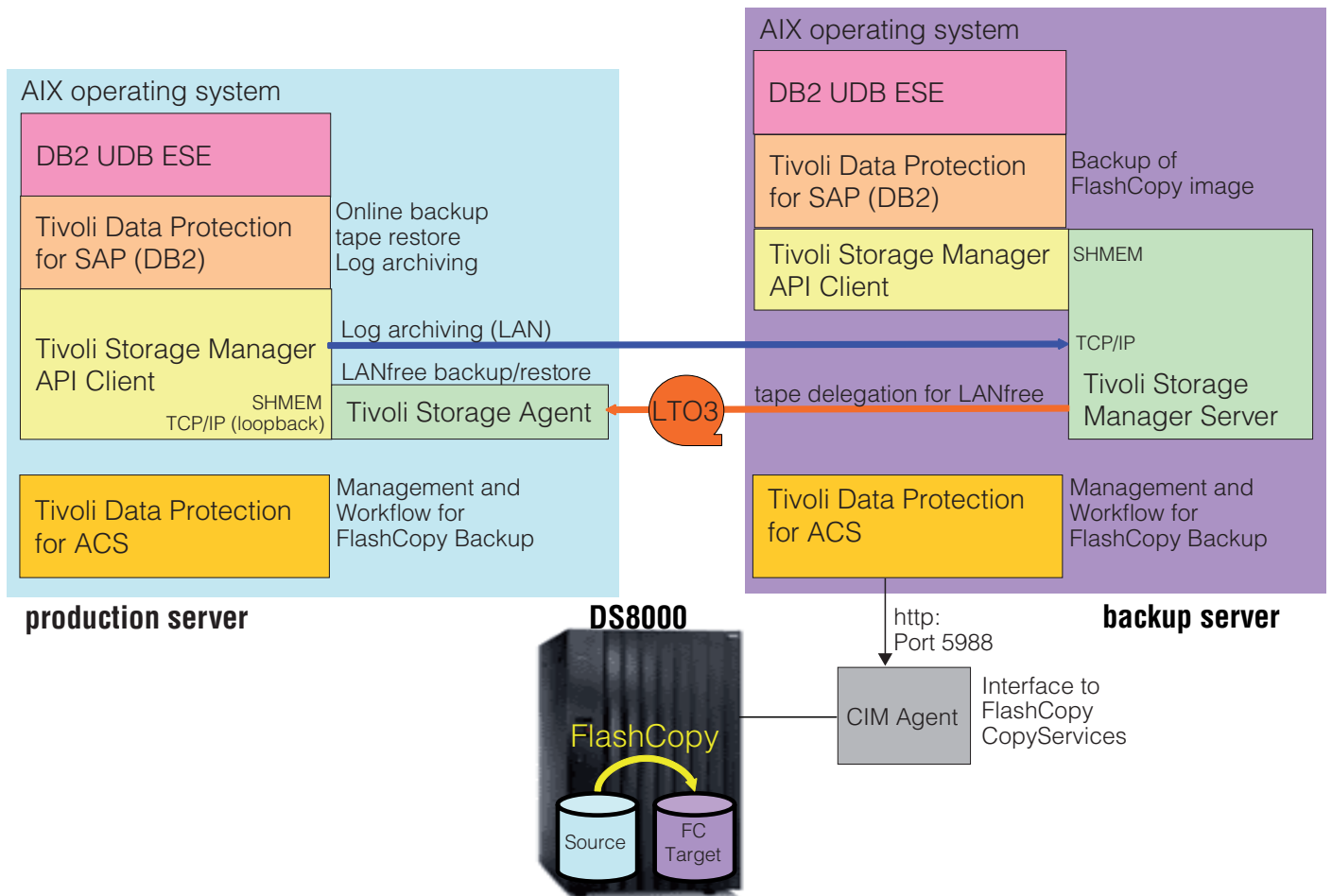
Figure 2 - 15: TSM implementation – component model

Proof of Design: Restore 60TB Database from Tape (KPI-2)

Figure 2 - 16 shows one stage of the disaster recovery KPI, a full restore of the 60TB database from tape. For a restore, partition 0 must be restored first and therefore two tape drives are used for this step with a throughput of 700GB per hour.

The remaining 32 partitions are restored in parallel, using one tape drive each with an individual throughput of 470GB per hour. The total restore time is 6 hours, 40 minutes.

Prior to the restore, a full backup was completed using the same configuration. The backup was achieved in 5hours, 27 minutes.

For KPI-2, a final step is required: the roll-forward of 2TB of logs, simulating the recovery of several days' work at the DR site. The roll-forward is shown in Chart 1 5 below. The recovery time is dominated by partition 0; the remaining 32 partitions roll-forward in parallel.

With a non-parallel database, the recovery time would be the sum total of all 33 individual recovery times (the time for partition 0 + (32 * the individual parallel partition time)). Instead, using the shared nothing parallel database design, the roll-forward of 2TB is completed in 2hours 52 minutes.
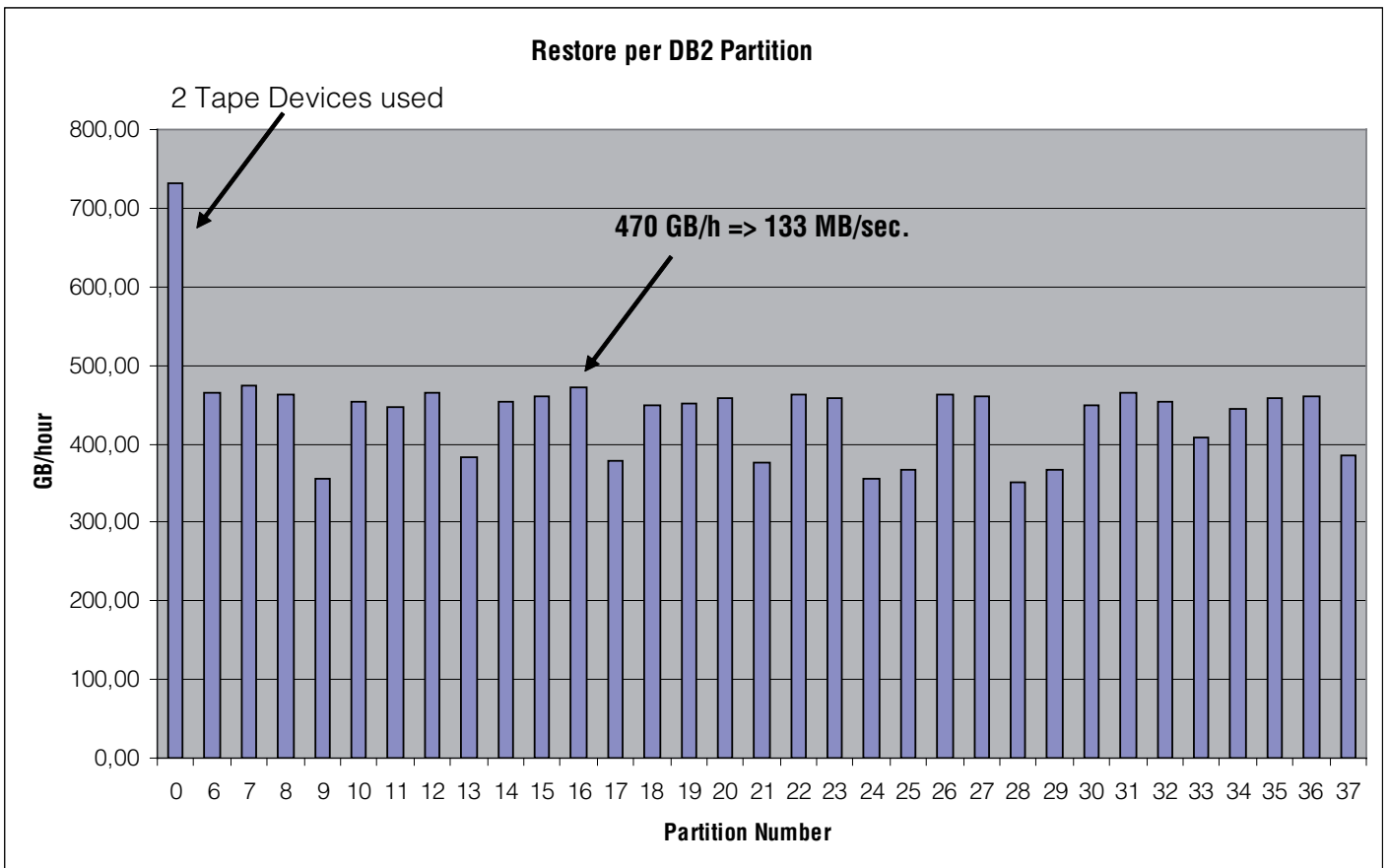
Figure 2 - 16: Restore from tape – 60TB
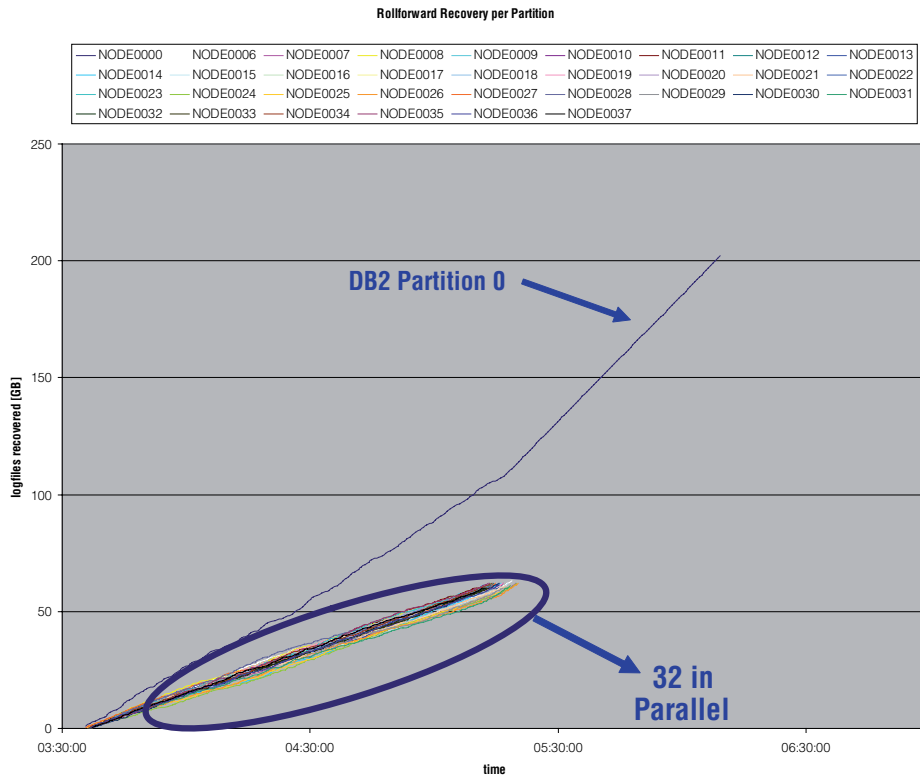
Figure 2 - 17: Disaster recovery roll-forward of 2TB of logs

Results of administration KPIs

The graph below shows the results of the administration KPIs and proof of design. The results of the proof of concept show that using parallel technology; extremely large databases can be administered effectively within a surprisingly small maintenance window. DB2 DPF shows the strength of its parallel design, and the Tivoli Storage Manager tools, Tivoli Storage Manager Server and Tivoli Storage Manager for ACS agents, demonstrate how they can manage multiple levels of parallelization: parallel database, multiple database hosts, multiple storage servers, and multiple target devices.

Ref.	Description	Requirement	Achievement
KPI-1	Recovery of an entire days' work Restore database and roll forward 500GB of logs	<8hrs	2hr 40min using IBM FlashCopy
KPI-2	Disaster recovery Complete database restore from tape plus roll forward of 2TB of logs	<24hr	10hrs Using LTO3 tapes
KPI-3.a	Daily backup to tape Full database backup to tape	<12hr	6hr 10min using LTO3
KPI-3.b	Daily backup to disk Completed incremental FlashCopy of production day	<8hr	30min DS8300 Turbo

Figure 2 - 18: Administration KPI achievements

P5 595 server design

The resource distribution and logical partition design of the p595 was completed in several steps. The evolution of the server design was determined by the proof of concept starting point, the DB2 evolution, and the scaling requirements. The basic plan is depicted in Figure 2-19.

The Nestlé clone was delivered on AIX 5.2 and DB2 V8. The objective was to introduce the newest technology, paving a migration path, and showing proof of benefit. Phase 1 introduced AIX 5.3 and the benefits of the hardware multi-threading (SMT)

along with the new DB2 V9, with scalability and performance benefits as well as new technology such as data compression. Phase 2 introduced micro-partitions and processor sharing.

To determine the initial design for the application servers, and begin the capacity planning for Phase 2, a series of load profiling tests were done. These provided the first "rule of thumb" for resource distribution. The graph in Figure 2-20 shows the DATABASE:APPS ratio done for the individual load profiles and the combined load for KPIs A and D at 20TB.
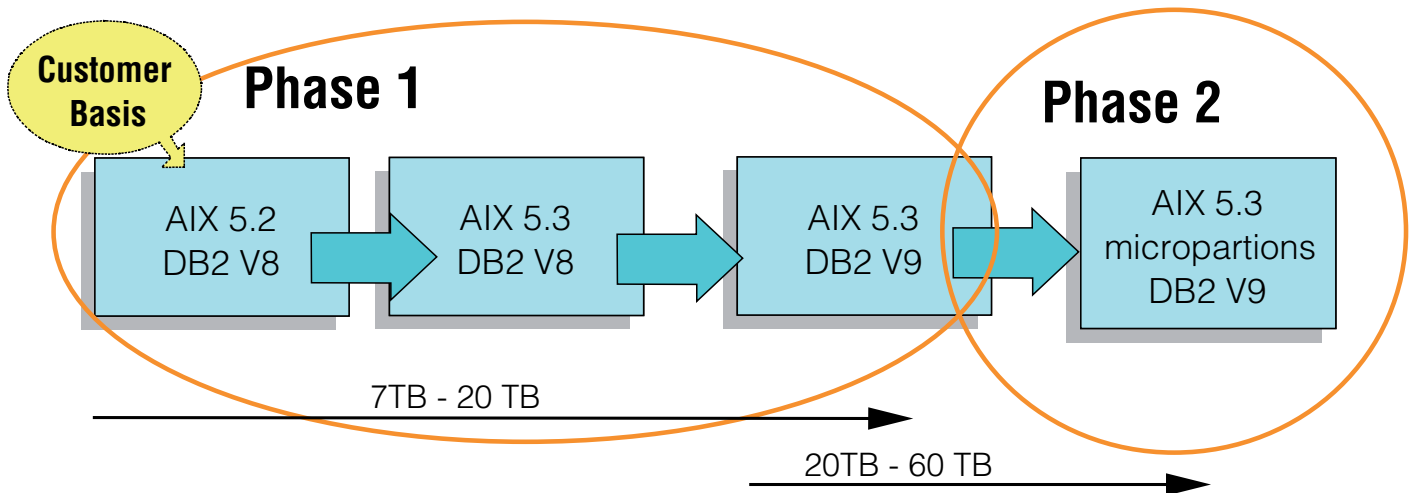
Figure 2 -19: Evolution plan
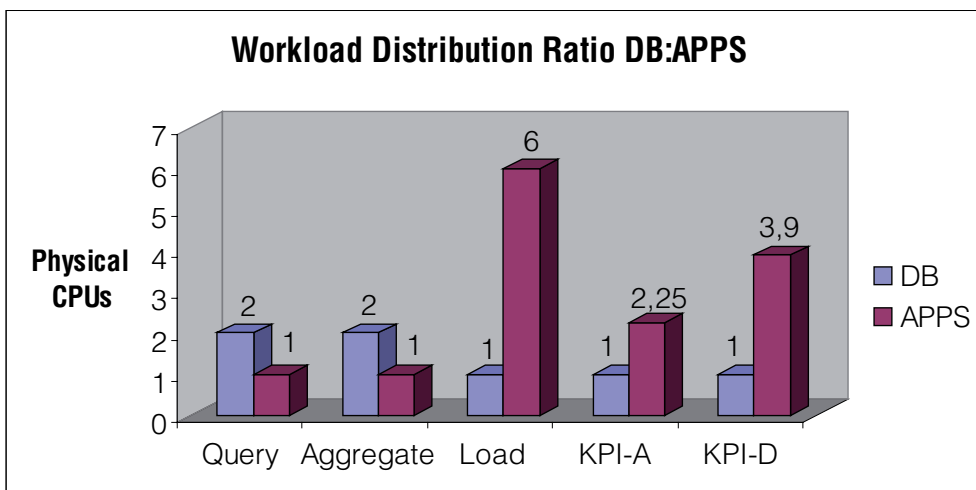
Figure 2 - 20: Database to application server ratios

Virtualization design for shared processor pool

Figure 2 - 21 shows the virtualization design for the final implementation at 60TB. The five database LPARs are distributed across the five p595 servers, one per machine. The first p595 is dedicated to administration-type work and load drivers. It contains the DB2 partition 0, the SAP software central instance, the data aggregation driver, and the data load extractors.

There is an online partition on each system to support the reporting. The HTTP-based online reporting load is distributed via an SAP load balancing tool, SAP Web Dispatcher®.

A Tivoli Storage Manager for ACS agent is placed on each of the p595 servers to manage the parallel LANFREE backup. Using the shared processor technology of p595 servers, a priority schema was implemented which gave highest priority to the database, medium priority to the online, and lowest priority to all batch activity. The storage agents were given a fixed CPU entitlement to ensure compliance with the SLA.

This entitlement can be used by other workloads when no backup is active. The highly application server CPU-intensive data loading scenario was given very low priority, but allowed to access all resources not used by higher priority systems. This allowed these application servers to be driven to full capacity while not impacting the response time KPIs of reporting.
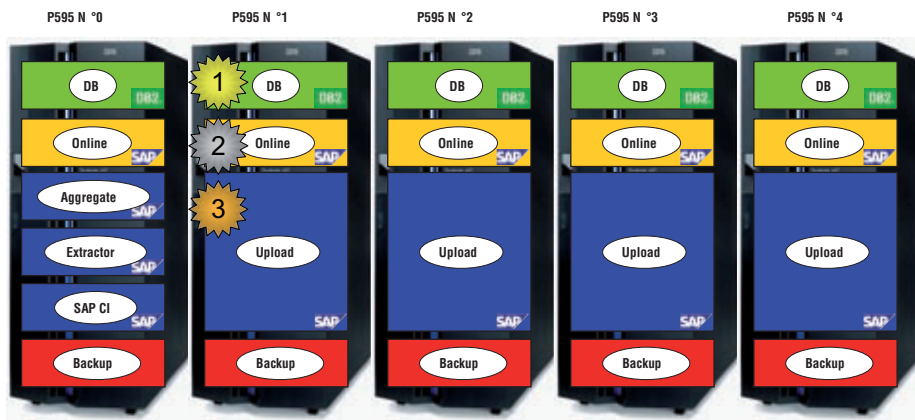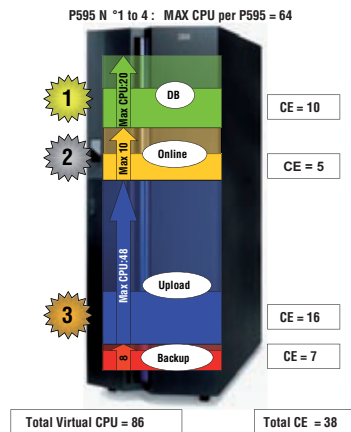

Figure 2 - 21: 60 TB server design

Figure 2 - 22: Virtualization Priority Schema

Infrastructure scaling

Figure 2 - 23 depicts the hardware scaling on the p5 595 as the load scenario was scaled by a factor of five times the baseline load. The scaling diagram separates the application servers from the database to show the load effects on the different components.

Figure 2 - 24 shows the physical resource scaling in relation to the load increase: for a load scaling of factor five, the physical resource increase was only factor of 3.5. In Figure 2 - 24 the physical CPUs have been normalized to relative GHz to factor in the speed increase from 1.9GHz to 2.3GHz. Between KPID at 20TB and KPID at 60TB, the system was moved into the final distributed configuration. Here the increase is related to the same load but for a much larger database. The final point is five times the baseline load at 6TB. The increase in application server requirements is factor 3.8, and the scaling factor for the database is 2.8 for a factor five load increase. The database design scales very efficiently.
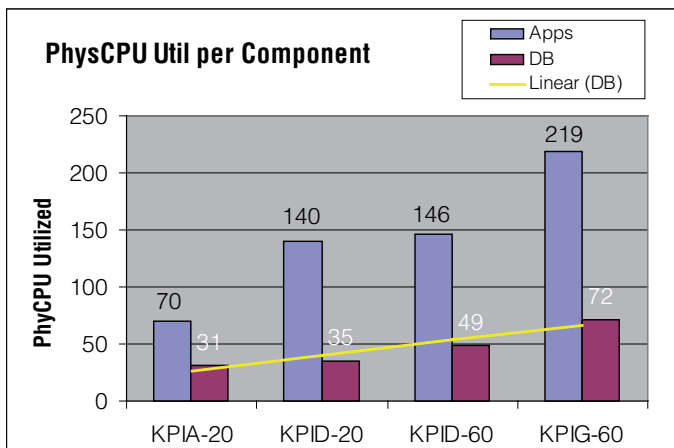
Figure 2 - 23: Total physical CPUs utilized
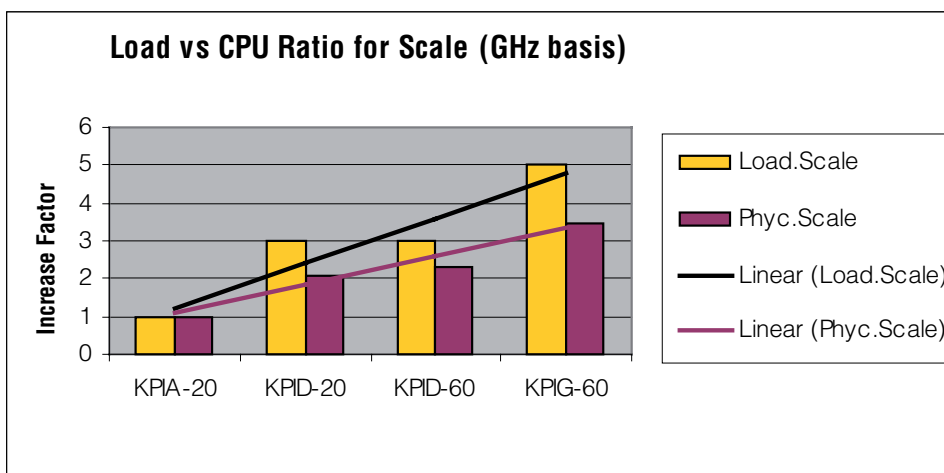
Figure 2 - 24: Resource scaling vs. load scaling

Online KPI achievements

Figure 2 - 25 follows the load throughput over the landscape evolution until achieving the final high-load combined KPI represented by KPI-G: 125 million records/hr in loading, 75 million records/hr in aggregation, with concurrent reporting rate of 2.08/sec, at 60TB.

The steps through the hardware were completed primarily using the KPI-D load requirements. In the final runs, the parallel landscape, implemented in SPLPARs (shared processor LPARs)

was used to scale the load to meet the KPI-G requirements.

The online achievements were the result of the p5 p595 server flexibility in handling diverse concurrent load requirement, the DB2 parallel database functionality, which supports a very broad based scaling of the database, and the dedicated "shared nothing" storage design. The project ended with the high-end KPI achievements for Nestlé, and there was still considerable scalability potential in this solution.
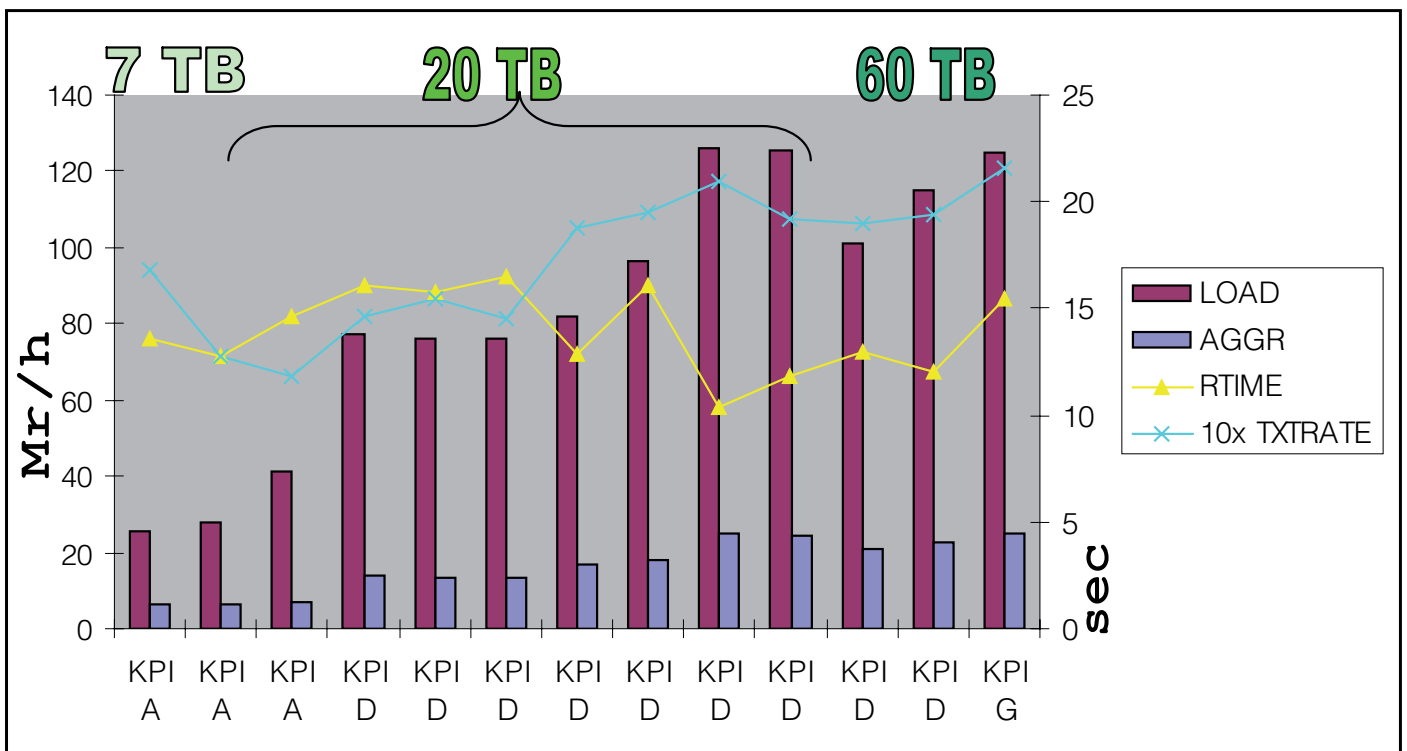
Figure 2 - 25: Evolution of the online KPI tuning and scale-up throughout the proof of concept

"The combination of SAP NetWeaver® BI and IBM infrastructure components show an unbeatable scalability addressing and fulfilling the requirements at high-end BI scenarios in the industry today."

Dr. Thomas Becker
SAP NetWeaver Business Intelligence Development Manager
SAP AG

Project achievements

The initiative paved the way for Nestlé's future business intelligence architecture implementation worldwide, and provided a very high-end proof point for best practices and directions for such systems in general.

Performance & throughput

A proven architectural design, based on IBM & SAP technologies, which can support Nestlé requirements for the next 2 to 3 years time (technology scalability proof-point).

Flexibility

IBM System p virtualization features enable the architecture to support a very broad mix of workloads. This option both capitalizes on Nestlé investments (skills and technology) and resolves the problems experienced by Nestlé caused by large variations in volumes and type of workload. Virtualization makes it possible to prioritize sensitive load types, such as the reporting queries, while utilizing the full capacity of the available resources.

Manageability

Proven and extendable architecture, based on parallelization, which allows an extremely large database to be managed well within the maintenance window allowed by the business. This design covers the spectrum from "business as usual" maintenance to full disaster recovery.

IBM

For more information:

To learn more about the solutions from IBM
and SAP visit: **ibm-sap.com**

For more information about SAP products and
services, contact an SAP representative or visit:
sap.com

For more information about IBM products and
services, contact an IBM representative or visit:
ibm.com

Contacts:

IBM

    Francois Briant, fbriant@fr.ibm.com

    Carol Davis, Davis@de.ibm.com

    Thomas Rech, Thomas.Rech@de.ibm.com

    Herve Sabrie, herve_sabrie@fr.ibm.com

For further questions please contact the
IBM SAP International Competency Center via
issic@de.ibm.com